UCRL-ID-137883

# Heterodyne Mask Blank Defect Detection LDRD Project Final Report

*M. A. Johnson, G. E. Sommargren*

*U.S. Department of Energy*

Lawrence Livermore National Laboratory

**February 4, 2000**

Approved for public release; further dissemination unlimited

## DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This work was performed under the auspices of the U. S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

This report has been reproduced directly from the best available copy.

Available electronically at http://www.doc.gov/bridge

Available for a processing fee to U.S. Department of Energy
And its contractors in paper from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-mail: reports@adonis.osti.gov

Available for the sale to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/ordering.htm

OR

Lawrence Livermore National Laboratory
Technical Information Department's Digital Library
http://www.llnl.gov/tid/Library.html

# Heterodyne Mask Blank Defect Detection
# LDRD Project Final Report

Michael A. Johnson
Gary E. Sommargren
February 4, 2000

## 1.0. Introduction

Mask blanks are the substrates that hold the master patterns for integrated circuits. Integrated circuits are semiconductor devices, such as microprocessors (μPs), dynamic random access memory (DRAMs), and application specific integrated circuits (ASICs) that are central to the computer, communication, and electronics industries. These devices are fabricated using a set of master patterns that are sequentially imaged onto light-sensitive coated silicon wafers and processed to form thin layers of insulating and conductive materials on top of the wafer. These materials form electrical paths and transistors that control the flow of electricity through the device.

For the past forty years the semiconductor industry has made phenomenal improvements in device functionality, compactness, speed, power, and cost. This progress is principally due to the exponential decrease in the minimum feature size of integrated circuits, which has been reduced by a factor of $\sqrt{2}$ every three years. Since 1992 the Semiconductor Industry Association (SIA) has coordinated the efforts of producing a technology roadmap for semiconductors. In the latest document, *The International Technology Roadmap for Semiconductors: 1999,* future technology nodes (minimum feature sizes) and targeted dates were specified and are summarized in Table 1.

**Table 1. Target dates for future technology nodes.**

| Year | 1999 | 2002 | 2005 | 2008 | 2011 | 2014 | 2017 |
|---|---|---|---|---|---|---|---|
| Technology node (nm) | 180 | 130 | 100 | 70 | 50 | 35 | 25 |

Lithography is the imaging technology for producing a de-magnified image of the mask on the wafer. A typical de-magnification factor is 4. Mask blank defects as small as one-eighth the equivalent minimum feature size are printable and may cause device failure. Defects might be the result of the surface preparation, such as polishing, or contamination due to handling or the environment. Table 2 shows the maximum tolerable defect sizes on the mask blank for each technology node. This downward trend puts a tremendous burden on mask fabrication, particularly in the area of defect detection and reduction.

**Table 2. Maximum tolerable sizes for mask blank defects.**

| Technology node (nm) | 180 | 130 | 100 | 70 | 50 | 35 | 25 |
|---|---|---|---|---|---|---|---|
| Max. defect size (nm) | 90 | 65 | 50 | 35 | 25 | 17 | 12 |

A new infrastructure for mask inspection will be required to keep pace with this aggressive roadmap. Depending on the specific lithography used for a particular generation, mask inspection specifics may change, but the methodology will essentially remain the same. Mask blanks will have to undergo 100% area inspection for defects larger than the maximum acceptable size. Since masks are becoming a significant cost factor in the ownership of lithography tools, this is a critical step—patterning defective mask blanks would be an economic disaster.

Inspection does not necessarily have to be done at the ultraviolet wavelength used for the lithography since defects at the mask blank level will interact with visible light, albeit very weakly. Techniques using visible light are appealing because they are familiar to the user, relatively straightforward to manufacture and safe to use, and when designed properly, extendable over many generations.

The technology used in commercial wafer inspection tools is currently the prime candidate for mask blank inspection. It is based on direct detection of scattered light from the defect in one or more directions. Figure 1 shows a typical setup with detectors in both the forward scatter direction (bright-field detection) and away from the specular direction (dark-field detection). In these setups the beam and/or mask blank is scanned to achieve full inspection of the blank. The scattered signal from a defect is therefore a short pulse immersed in the dynamic background scatter from the inherent surface roughness of the mask blank and in the light scattered from the optics and mechanical parts within the instrument. State-of-the-art instruments cannot detect defects smaller than 80 nm, insufficient for the next technology node.

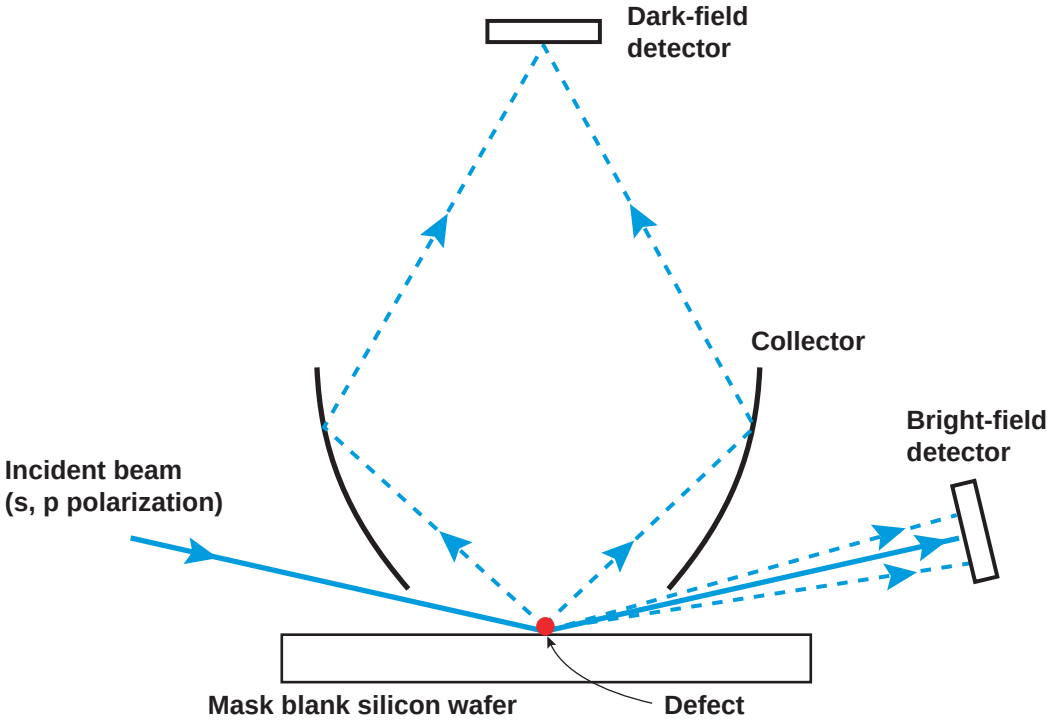


**Figure 1. Optical system in commercial wafer inspection tools based on measurement of scattered intensity.**

The research done over the last year addressed defect detection using a different approach—a heterodyne interference/synchronous detection technique that has the potential of

enhanced detection of the scattered light from small defects. This detection is accomplished by directly measuring the amplitude of the electric field of the scattered light using interference of the scattered light with a strong, frequency shifted, local oscillator beam. This technique could provide the basis for new visible light inspection equipment.

## 2.0. Heterodyne detection

Optical heterodyne detection—surprising as it may seem—predates[1] the invention of the laser. Furthermore, the general principles of heterodyne detection have a long history in radio frequency electronics.

In the optical system a square-law photodetector produces an electrical signal at the difference frequency of two optical waves combined on it. When one of these waves (the local oscillator, or LO) is made much stronger than the other (signal) wave, the sensitivity of the process can be much higher than for direct detection of the signal alone. In addition, the heterodyne detector has both strong frequency and strong directional selectivity; that is, it acts as both a receiver and an antenna. Careful alignment between the LO and signal beams is necessary to maintain a constant phase of the beat note across the face of the photodetector; therefore, heterodyne detection is most useful for detecting coherent, local sources. Many workers in many fields—laser spectroscopy, Doppler velocimetry, plasma diagnostics, profilometry, astronomy, LIDAR, to name a few—have taken advantage of these characteristics.

In the following sections we first review the principles of optical heterodyne detection and then continue with analysis of its application to the detection of very small particles lying on a rough surface. Section 2.1 describes the optical layout and electronic processing for a heterodyne system and reviews the fundamental properties of the system noise. Section 2.2 describes the optical scattering properties of a sub-wavelength sized conducting sphere and discusses the heterodyne signal to be expected when such a sphere is detected. Section 2.3 is a description of the fluctuating response of the heterodyne system to sub-wavelength-scale surface roughness. These fluctuations are a potential noise source for the system signal-to-noise ratio (SNR). Next, in Section 2.4 we bring together results of prior sections to derive an overall system SNR. Finally, in Section 2.5 we describe our experimental results with our early breadboard of a heterodyne system.

### 2.1. Heterodyne system description

**2.1.1. Optical layout.** Figure 2 shows the layout of an optical heterodyne system designed to detect small defects on “rough” surfaces. A single-frequency, single-spatial-mode laser is split at a beam splitter into two beams. One beam, called the probe, is focussed to a spot on the mask blank. The very well polished mask blank reflects most of the probe beam in a specular fashion, but the residual surface roughness scatters a few parts per billion (ppb) into a large solid angle. If the probe beam should happen to hit a defect such as a particle on the surface, then this diffuse scattering can be greatly increased.

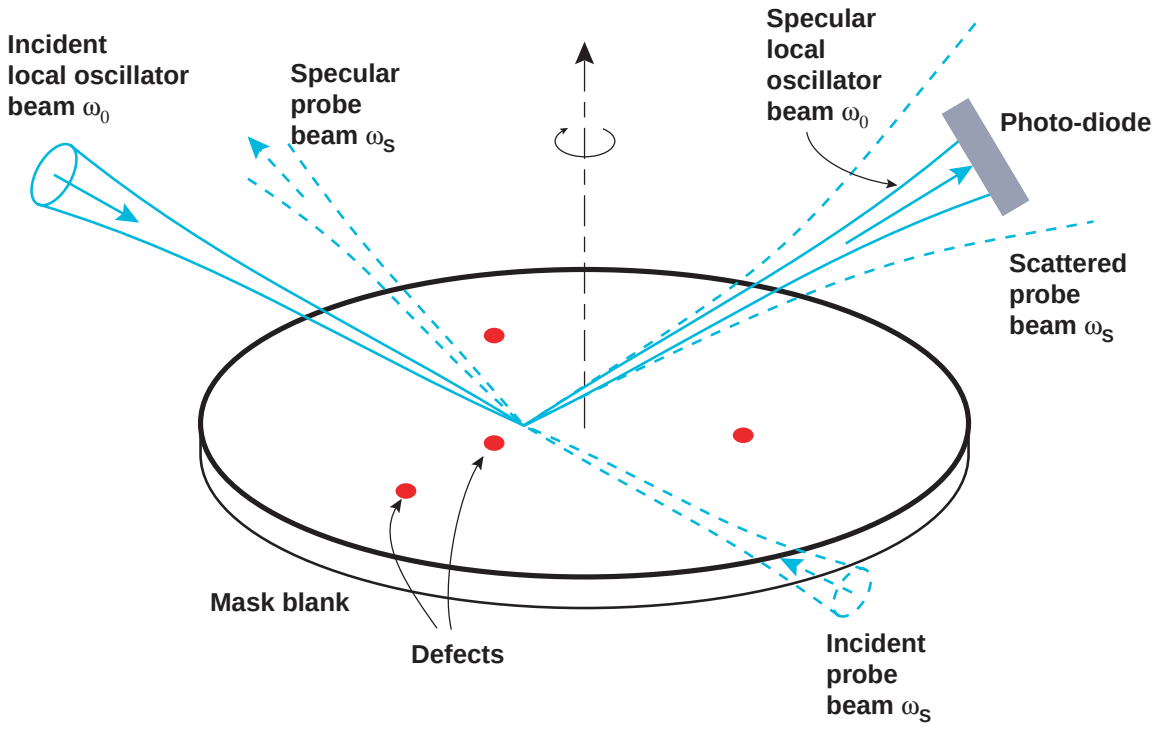


**Figure 2. Optical layout of a heterodyne system for defect detection.**

The second laser beam, called the local oscillator (LO) passes through an acousto-optic frequency shifter and is then focussed to the same spot on the mask blank as the probe laser, although its incident direction is made completely different from that of the probe. The LO beam also undergoes specular reflection (and a little diffuse scattering). A photo-diode is placed to collect all the specular reflection of the LO beam as well as the diffuse scatter from the probe beam that coincides with the specular LO beam.

It is the intensity beat between these two optical fields that constitutes the signal in the heterodyne system. The beat occurs at exactly the frequency used in the acoustic frequency shifter. In practice the mask blank is scanned rapidly through the focussed laser spot in order to find defects anywhere on the surface. The incident directions of the probe and LO lasers are designed so that there is no Doppler frequency shift in the photo-diode beat signal caused by the surface motion.

**2.1.2. Signal processing and electronics.** Figure 3 shows a schematic view of the signal path through the electronics of the system. At the top the optical fields of the scattered probe laser and the LO fall on the photo-diode, which in analogy with RF systems we call the photomixer. The photomixer output current passes through an RF bandpass filter centered at the beat frequency of the probe and the LO. This filter has a wide enough pass band to transmit amplitude variations in the beat signal as a defect passes through the laser spot on the mask blank.

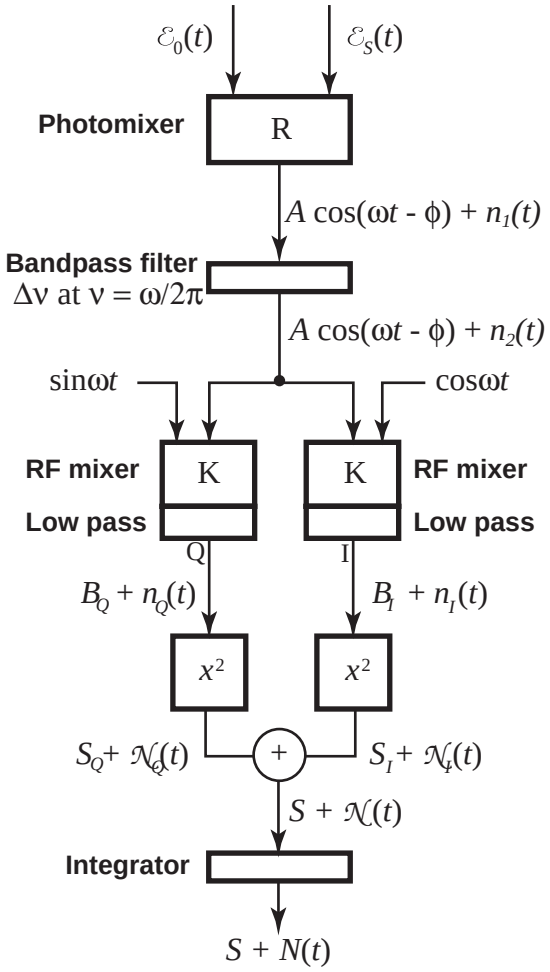


**Figure 3. Heterodyne system electronics configuration. The quantities labeled $n_1(t)$, $n_2(t)$, $n_Q(t)$, $\mathcal{N}_Q(t)$, etc. represent additive random noise.**

It is this beat signal amplitude that we want to measure, but we have no prior knowledge of its phase; therefore, we electronically split the beat signal into two inputs to RF mixers operating in quadrature. Each mixer is driven by an RF intermediate local oscillator (ILO) set to the known laser beat frequency. The two ILO's are 90 degrees out of phase. In this way one mixer demodulates the in-phase component of the photomixer current and the other mixer demodulates the quadrature-phase component. Low-pass filters following the RF mixers remove any sum-frequency components generated in the mixing process.

Finally, to recover the amplitude of the original beat signal and discard its phase the two outputs of the RF mixers are squared and summed. This final voltage is then averaged in an integrating circuit, typically with a time constant about equal to the time it takes a defect on the moving mask blank to pass through the laser spot.

**2.1.3. SNR for heterodyne system.** The following definitions apply to the quantities in Figure 3:

$\mathcal{E}_0(t) \equiv$ local oscillator field (real number)

$= \mathrm{Re}\left(E_0 e^{i\omega_0 t}\right)$

$\mathcal{E}_S(t) \equiv$ signal field (real)

$= \mathrm{Re}\left(E_S e^{i\omega_s t}\right)$

$\omega_0 \equiv$ local oscillator frequency

$\omega_S \equiv$ signal frequency

$\omega \equiv$ beat frequency

$= \omega_0 - \omega_S$

$I_0(t) \equiv$ local oscillator intensity (real)

$$= \frac{|E_0|^2}{2} \qquad [1]$$

$I_S(t) \equiv$ signal intensity (real)

$$= \frac{|E_S|^2}{2} \qquad [2]$$

$I(t) \equiv$ total intensity (real)

$$= I_0 + I_S + \tfrac{1}{2}\left[E_0 E_S^* e^{i\omega t} + c.c.\right] \qquad [3]$$

The first term in the total intensity is the LO intensity, which gives rise to a DC current in the photomixer:

$$i_{DC} = R I_0 \quad , \qquad [4]$$

where $$R \equiv \frac{\eta q \mathcal{A}}{h\nu} \qquad [5]$$

and

$\eta \equiv$ photomixer quantum efficiency

$q \equiv$ charge on the electron

$\mathcal{A} \equiv$ area of photomixer

$h\nu \equiv$ energy of a photon

The third term is the heterodyne mixing term, which gives rise to an AC current

$$i_{AC} = A\cos(\omega t - \phi) = \tfrac{1}{2} R\left[E_0 E_S^* e^{i\omega t} + c.c.\right] \quad . \qquad [6]$$

Thus the amplitude $A$ of the sinusoidal current component is

$$A = 2R\sqrt{I_0 I_S} \quad . \qquad [7]$$

The bandpass filter following the photomixer passes the signal with unity gain; the filter affects only the noise.

The pair of matched RF mixers operate in quadrature to produce the DC signals given by (bar designates time average)

$$B_I = K\,\overline{A\cos(\omega t-\phi)\cos\omega t} = \frac{KA}{2}\cos\phi$$
$$B_Q = K\,\overline{A\cos(\omega t-\phi)\sin\omega t} = \frac{KA}{2}\sin\phi \qquad [8]$$

The RF squaring circuits produce the outputs

$$S_I = B_I^2 = \left(\frac{KA}{2}\right)^2\cos^2\phi$$
$$S_Q = B_Q^2 = \left(\frac{KA}{2}\right)^2\sin^2\phi \qquad [9]$$

Finally the system signal output voltage is

$$S = S_I + S_Q$$
$$= (KR)^2 I_0\, I_S \qquad [10]$$

Now we are ready to treat the noise in the heterodyne system. We assume that all the electronic components are ideal and produce no noise of their own. This is justified because we can make the shot-noise fluctuations in the DC photomixer current larger than any other fixed noise sources if we make the DC current large enough.

First, a word about definitions of noise: by system noise we mean the noise that exists in the absence of any signal, that is, the noise that would contribute to false positive detection. One can also ask about the error in measuring a true signal. In this case the interaction of the signal and the noise together are important. We are not showing this latter analysis because our real interest is in pushing the limits of defect detectability.

In the absence of any signal from the probe beam, the only current in the photomixer comes from the LO, as shown in Eq. [4]. The current fluctuations are due to the quantized charge of the current carriers. At the output of the bandpass filter following the photomixer these current fluctuations, expressed as a variance, are given by[2]

$$\left\langle n_2^2(t)\right\rangle \equiv \left\langle \left(i-i_{DC}\right)^2\right\rangle = 2q\,i_{DC}\Delta\nu\ , \qquad [11]$$

where the angle brackets denote a statistical average. Referring to Fig. 3, this noise can also be written as a nearly sinusoidal current given by

$$n_2(t) = n_\omega(t)\cos\left[\omega t+\alpha(t)\right] \qquad [12]$$

where $n_\omega(t)$ and $\alpha(t)$ are random variables and $\alpha(t)$ has a uniform random distribution over the range 0 to $2\pi$.

From Eqs. [11] and [12] we have

$$\left\langle n_\omega^2(t)\right\rangle = 4q\ i_{DC}\,\Delta\nu\ . \qquad [13]$$

Next, the quasi-sinusoidal noise voltage $n_2(t)$ passes through the two RF mixers and their subsequent low pass filters to produce

$$\begin{aligned} n_I(t) &= K\, n_\omega(t)\overline{\cos[\omega t + \alpha(t)]\cos\omega t} \\ &= \tfrac{1}{2} K\, n_\omega(t)\cos[\alpha(t)] \\ n_Q(t) &= \tfrac{1}{2} K\, n_\omega(t)\sin[\alpha(t)] \end{aligned} \qquad [14]$$

The RF mixers have preserved the flat spectrum of the noise, but these spectra have been translated down to DC and folded over; that is, the spectrum of $n_Q(t)$ and $n_I(t)$ now extend from DC to a frequency of $\Delta\nu/2$. We assume that the circuitry removes the DC components without significantly affecting this noise bandwidth.

Next we put these outputs of the RF mixers through RF squaring circuits to produce the DC voltages

$$\begin{aligned} \mathcal{N}_I(t) &= n_I^2(t) \\ \mathcal{N}_Q(t) &= n_Q^2(t) \end{aligned} \qquad [15]$$

It can be shown[3] that the fluctuations in these voltages are

$$\begin{aligned} \delta\mathcal{N}_I^2 &\equiv \left\langle \left(\mathcal{N}_I - \langle \mathcal{N}_I \rangle\right)^2 \right\rangle \\ &= 2\left\langle n_I^2(t)\right\rangle^2 \\ \delta\mathcal{N}_Q^2 &= 2\left\langle n_Q^2(t)\right\rangle^2 \end{aligned} \qquad [16]$$

The summing circuit at the bottom of Fig. 3 produces a mean square system noise of

$$\delta\mathcal{N}^2 = \delta\mathcal{N}_I^2 + \delta\mathcal{N}_Q^2 \quad . \qquad [17]$$

Recall that this noise has a bandwidth of $\Delta\nu/2$. We assume that the heterodyne measurement is integrated for a time $T$, where

$$\frac{1}{2T} = \frac{\Delta\nu}{2} \qquad [18]$$

Using the results of Eqs. [13]–[18] and Eqs. [4] and [5] we have

$$\delta\mathcal{N}^2 = \left(\frac{K^2 q\, R\, I_0}{T}\right)^2 \quad . \qquad [19]$$

Finally, from Eqs. [10] and [19] we can define a system signal-to-noise ratio assuming that the photomixer shot noise is the only noise:

$$\begin{aligned} \mathrm{SNR} &\equiv \frac{S}{\left(\delta\mathcal{N}^2\right)^{1/2}} = \frac{R}{q} I_S\, T \\ &= \frac{\eta I_S \mathcal{A}}{h\nu} T \end{aligned} \qquad [20]$$

We can make this look particularly simple by defining an equivalent input "noise" optical intensity, meaning an intensity that will have an SNR of unity in a measuring time $T$.

$$I_{\substack{het \\ noise}} \equiv \frac{h\nu}{\eta \mathcal{A} T} \qquad [21]$$

which is the intensity causing one absorbed photon in an area $\mathcal{A}$ in a time $T$. The SNR therefore becomes

$$\mathrm{SNR} = \left( \frac{I_S}{I_{\substack{het \\ noise}}} \right) . \qquad [22]$$

Equations [21] and [22] show several of the striking and well-known characteristics of optical heterodyne detection. First, the SNR does not depend on the LO laser power; the signal and the noise increase together with increased LO. In practical terms, this means that the LO power can usually be increased until Eq. [21] represents the dominant source of noise. Another characteristic of heterodyne systems is that the equivalent input noise corresponds to a single detected photon arriving during the measurement time and occupying a spatial mode matched to the LO. This input noise has many of the same characteristics as the well-known "zero-point fluctuations" found in descriptions of other quantum phenomena, such as the spontaneous emission of laser amplifiers[4].

## 2.2. Signal from a point scatterer

We next calculate the heterodyne signal that we would expect from a small defect on the mask blank illuminated by the probe and LO beams. The semiconductor industry follows a standard of calibrating all defects in terms of the size of a small sphere that scatters the same amount as the defect. For spheres that are much smaller than an optical wavelength this scattering can be calculated from knowledge of the dielectric properties of the sphere and the surface, from the optical wavelength, and from the diameter of the sphere.

**2.2.1. Small-particle scattering.** There is a large body of literature describing the electromagnetic scattering from sub-wavelength sized spheres, much of it coming under the name of Mie theory[5]. For an isolated sphere one can show from dimensional arguments that the fractional scattering has the form

$$\frac{P_S}{P_L} \propto \frac{D^6}{\lambda^4 d^2} \quad , \qquad [23]$$

where

$P_S \equiv$ total scattered laser power
$P_L \equiv$ incident laser power
$D \equiv$ diameter of scattering particle
$\lambda \equiv$ wavelength of scattered laser
$d \equiv$ diameter of laser spot

The dimensional argument goes something like this:

1) $d^{-2}$: The scattered power is proportional to the incident intensity at the particle; the incident intensity $\propto d^{-2}$.

2) $\lambda^{-4}$: The total power radiated by an accelerated charge is proportional to the square of its acceleration. The acceleration of a sinusoidally driven charge is proportional to the driving frequency squared times the amplitude of motion. Far below resonance the amplitude of motion of such a charge is independent of frequency. Hence, the total power radiated is proportional to the fourth power of the driving frequency, or $\lambda^{-4}$.
3) $D^6$: The sphere is much smaller than a wavelength, so all the microscopic radiators it contains contribute in phase to the radiated field. The strength of the radiated field is proportional to the number of radiators in the sphere, which is proportional to its volume; that is, the optical polarizability of the sphere is proportional to its volume. The radiated power is proportional to the square of the radiated field, hence to $D^6$.

The scattering properties of a small sphere sitting on a surface are much more complicated than for an isolated sphere and the problem is best treated numerically. We have found, however, that for large angles of incidence Eq. [23] applies well, particularly in its dependence on sphere diameter *D*. For example, numerical simulation of a Si sphere sitting on a Si surface gives the following approximate formula for P-polarized light at 70° angle of incidence:

$$\frac{P_S}{P_L} \approx 1000 \frac{D^6}{\lambda^4 d^2} \quad , \qquad [24]$$

which fits the numerical simulation well for a range of *D* from 5 nm to 50 nm at wavelengths near 500 nm.

**2.2.2. Overlap of scattered wave with LO.** For our purposes here, it is adequate to assume that when the spherical particle is illuminated by the probe laser the resulting scattered wave has a spherical wave front centered on the particle. We can also assume that this wave has an amplitude that varies smoothly with scattering angle. We will ignore the effects of optical polarization, as if the scattered wave and the LO have polarizations that are parallel. This is clearly not possible for all directions of probe and LO, but the added complexity is beyond our scope at this time. Thus our problem reduces to calculation of the beat signal in the integrated intensity across the photomixer of the combined spherical wave from the particle and the optical field from the LO.

We need to generalize the form of the amplitude *A* of the sinusoidal component of the photomixer current as shown in Eq. [7] because we cannot assume that the intensities of the scattered wave and the LO are constant across the face of the photomixer. For varying intensities we have

$$A = R \left| \iint E_0(x,y,z) E_S^*(x,y,z)\, dx\, dy \right| \qquad [25]$$

where the integral is taken over the surface of the photomixer. We will assume that the particle is located at the point $(x_S, y_S, z_S)$ and that the complex signal field is a spherical wave emanating from that point:

$$E_S(x,y,z) = \sqrt{2P_S}\, \frac{e^{-ikr_S}}{r_S} \cos(\hat{z}, \vec{r}_S) \quad , \qquad [26]$$

where $P_S$ is the optical power per unit solid angle in the forward (+$z$) direction, where $\vec{r}_S$ is a vector from $(x_S, y_S, z_S)$ to $(x,y,z)$, and where

$$r_S = |\vec{r}_S| = \left[(x - x_S)^2 + (y - y_S)^2 + (z - z_S)^2\right]^{1/2}. \qquad [27]$$

The $\cos(\hat{z}, \vec{r}_S)$ factor in Eq. [26] means that the field peaks in the +$z$ direction and falls to zero in the $x$-$y$ plane. In fact we assume that $E_S(x,y,z)$ is zero for negative $z$. Thus Eq. [25] becomes

$$A = R\sqrt{2P_S}\left| \iint_{-\infty}^{\infty} E_0(x,y,z)\, \frac{e^{ikr_S}}{r_S} \cos(\hat{z}, \vec{r}_S)\, dx\, dy \right| . \qquad [28]$$

We have extended the integral to infinite limits to signify that the surface of the photomixer captures all of the LO field $E_0(x,y,z)$.

The integral in Eq. [28] is a well-known form from the theory of scalar diffraction. Using a Green's function approach to the solution of the scalar wave equation results in the Rayleigh-Sommerfeld equation[6]

$$E_0(x_S, y_S, z_S) = \frac{1}{i\lambda} \iint_{-\infty}^{\infty} E_0(x,y,z) \frac{e^{ikr_S}}{r_S} \cos(\hat{z}, \vec{r}_S)\, dx\, dy \quad , \qquad [29]$$

which describes how an optical field $E_0(x,y,z)$ known over a surface propagates to an observation point $(x_S, y_S, z_S)$. In terms of the heterodyne problem at hand, this means that the overlap integral at the photomixer of the LO field and a spherical wave is related to the strength of the LO field evaluated at the position of the scattering particle.

Without needing to know the explicit form of the LO optical field we can therefore use Eq. [29] to evaluate the integral in Eq. [28]:

$$A = \lambda R\sqrt{2P_S}\, |E_0(x_S, y_S, z_S)| \quad . \qquad [30]$$

This very general result says the following about the integrated beat note between an arbitrary local oscillator and a point emitter producing a spherical wave. If you move the point emitter around in three dimensions, keeping the emitted power constant, then the amplitude of the heterodyne beat note exactly tracks the amplitude of the LO field at the location of the point. This is true regardless of where the detector is located, as long as the detector collects all the overlapping light. In practice the photomixer must just collect all of the LO beam, since it is much more localized than the scattered probe beam.

We showed in Eqs. [10] and [7] that the signal $S$ at the output of the heterodyne system is proportional to the quantity $A^2$, where $A$ is given by Eq. [30]. Thus the heterodyne signal is simply proportional to the product of the LO intensity and the probe laser intensity, both evaluated at the position of the scattering particle. It is not necessary to evaluate the integral of the overlap of the two optical fields at the photomixer surface as long as the photomixer collects the entire LO beam.

Finally, another consequence of Eq. [30] is that the overlap integral in Eq. [28] is independent of where the integral is done, as long as all the LO power is included. That is, we can evaluate the integral in any plane we choose and we will get the same result as if we cal-

culated the integral over the photomixer surface. We will make use of this invariance in the following section.

### 2.3. Heterodyne response from a rough surface

Even without any defects on the mask blank surface, the slightly rough surface of the blank scatters some of the probe laser onto the photomixer, producing a heterodyne beat note. This roughness is, of course, not very great. In fact, the blanks are super-polished to the point that only a few parts per billion (ppb) of the probe laser are scattered diffusely. In this section we will derive the relationship between this surface roughness and the heterodyne "signal" it produces.

We first assume that the optical fields of both the probe beam and the LO have flat wave fronts near the mask blank surface and furthermore that the directions of propagation are characterized by the direction cosines $(\alpha_S, \beta_S, \gamma_S)$ for the probe and $(\alpha_0, \beta_0, \gamma_0)$ for the LO. The *z*-axis coincides with the surface normal of the blank. We also assume that the surface roughness is described by the height vs. position given by $h(x,y)$, where the average value of $h$ is zero and its probability distribution is gaussian.

We can then use Eq. [25] to write

$$\left\langle A^2 \right\rangle = R^2 \left\langle \left| \int_{-\infty}^{\infty} \int E_0(x,y) E_S^*(x,y) e^{ik[x\,\delta\alpha + y\,\delta\beta]} e^{ik\,\delta\gamma\,h(x,y)} dx\,dy \right|^2 \right\rangle \quad , \quad [31]$$

where the angle brackets denote a statistical average over the properties of the rough surface and where $\delta\alpha \equiv \alpha_S - \alpha_0$, $\delta\beta \equiv \beta_S - \beta_0$ and $\delta\gamma \equiv \gamma_S - \gamma_0$. We then define a phase difference

$$\phi(x,y) \equiv k\,\delta\gamma\,h(x,y) = \frac{2\pi\,\delta\gamma\,h(x,y)}{\lambda} \quad . \quad [32]$$

Expanding the squared magnitude we have

$$\left\langle A^2 \right\rangle = R^2 \int_{-\infty}^{\infty} \int \int_{-\infty}^{\infty} \int E_0(x_1,y_1) E_S^*(x_1,y_1) E_0^*(x_2,y_2) E_S(x_2,y_2) \times e^{ik\left[(x_1-x_2)\delta\alpha + (y_1-y_2)\delta\beta\right]} \left\langle e^{i\left[\phi(x_1,y_1) - \phi(x_2,y_2)\right]} \right\rangle dx_1\,dy_1\,dx_2\,dy_2 \quad [33]$$

The statistical average in the integrand of Eq. [33] can be evaluated in terms of the spatial correlation function of the phase shift $\phi$, or equivalently, of the surface height $h(x,y)$. Using the properties of the characteristic function of a gaussian random variable, one can show[7] that

$$\left\langle e^{i\left[\phi(x_1,y_1) - \phi(x_2,y_2)\right]} \right\rangle = \exp\left\{-\sigma^2\left[1 - \rho(\varepsilon,\tau)\right]\right\} \quad , \quad [34]$$

where the variance $\sigma^2$ of the phase is given by

$$\sigma^2 \equiv \left\langle \phi^2(x,y) \right\rangle \quad [35]$$

and is independent of position. We have defined the differences

$$\begin{aligned} \varepsilon &\equiv x_1 - x_2 \\ \tau &\equiv y_1 - y_2 \end{aligned} \quad . \quad [36]$$

We also use the normalized correlation function ρ of the phase as given by

$$\sigma^2\rho(\varepsilon,\tau) \equiv \langle \phi(x,y)\phi(x+\varepsilon,y+\tau)\rangle \quad , \qquad [37]$$

where ρ is independent of the position $(x,y)$ and does not depend on the signs of ε or τ. Furthermore, we assume that the surface heights vary by much less than an optical wavelength, so that $\sigma^2 << 1$. Using the property of the correlation function that $|\rho| \le 1$, we can expand the exponential in Eq. [34] to get

$$\left\langle e^{i\left[\phi(x_1,y_1)-\phi(x_2,y_2)\right]}\right\rangle \approx 1+\sigma^2\rho(\varepsilon,\tau) \ . \qquad [38]$$

Next, we assume that the phase φ remains spatially correlated only over a distance that is much shorter than the characteristic distance over which the LO and probe field amplitudes $E_0$ and $E_S$ change. We therefore define the new variables

$$\begin{aligned} x &\equiv \frac{x_1+x_2}{2} \\ y &\equiv \frac{y_1+y_2}{2} \end{aligned} \qquad [39]$$

and use them in Eq. [33] to get

$$\begin{aligned} \left\langle A^2\right\rangle = R^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} &|E_0(x,y)|^2 |E_S(x,y)|^2 \\ &\times e^{ik[\varepsilon\delta\alpha+\tau\delta\beta]}\left[1+\sigma^2\rho(\varepsilon,\tau)\right]dx\,dy\,d\varepsilon\,d\tau \end{aligned} \qquad [40]$$

The integral can now be expanded and separated to give

$$\begin{aligned} \left\langle A^2\right\rangle = R^2 G \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ik[\varepsilon\delta\alpha+\tau\delta\beta]}d\varepsilon\,d\tau \\ +\sigma^2 R^2 G \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ik[\varepsilon\delta\alpha+\tau\delta\beta]}\rho(\varepsilon,\tau)d\varepsilon\,d\tau \end{aligned} \qquad [41]$$

where

$$\begin{aligned} G &\equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |E_0(x,y)|^2 |E_S(x,y)|^2\, dx\,dy \\ &= 4\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_0(x,y)\, I_S(x,y) dx\,dy \end{aligned} \qquad [42]$$

The first integral in Eq. [41] is the result we would obtain if the surface were perfectly flat, that is, if $h(x,y) \equiv 0$. Thus it is the heterodyne beat signal we would have between the probe and the LO if they simply crossed at the surface and continued with no perturbation. We can make this signal as small as we like by making sure the probe and LO are physically separate at some point. That is, if there is some part of the system where the integral of the

overlap is zero, then that overlap remains zero as long as the two beams simply propagate in vacuum without perturbation.

The second integral in Eq. [41] looks like a Fourier transform of the normalized spatial correlation function ρ for the phase ϕ. This means that the integral can be interpreted as the power spectral density (PSD) of the phase ϕ. One can show[8] that if we define the PSD for the two-dimensional surface as

$$S(f_x, f_y) \equiv \left| \int_{-\infty}^{\infty} \int h(x,y) e^{2\pi i (f_x x + f_y y)} dx\, dy \right|^2 \quad , \qquad [43]$$

then the well-known Fourier transform relation between the PSD and the correlation function of the surface distribution $h(x,y)$ gives us

$$S(f_x, f_y) = \sigma^2 \left( \frac{\lambda}{2\pi\delta\gamma} \right)^2 \int_{-\infty}^{\infty} \int \rho(\varepsilon, \tau) e^{2\pi i (\varepsilon f_x + \tau f_y)} d\varepsilon\, d\tau \quad . \qquad [44]$$

Finally, we have from Eq. [41]

$$\langle A^2 \rangle = R^2 G \left( \frac{2\pi\delta\gamma}{\lambda} \right)^2 S\left( \frac{\delta\alpha}{\lambda}, \frac{\delta\beta}{\lambda} \right) \quad . \qquad [45]$$

To find a more specific form for the constant $G$ defined in Eq. [42], we will take the example of gaussian intensity profiles for both the probe and LO beams and we will also assume that the intensity profiles are adjusted to match at the mask blank surface. We find that

$$G = \frac{P_0 P_S}{\pi w_x w_y} \quad , \qquad [46]$$

where the total laser powers in the LO and probe are $P_0$ and $P_S$, respectively, and the waist radii at the $1/e^2$ intensity points are $w_x$ in the $x$ direction and $w_y$ in the $y$ direction. We can interpret $G$ as the product of the two powers divided by the area of the laser spot on the mask blank.

We know from Eqs. [10] and [7] that the average "signal" caused by the rough surface at the output of the heterodyne system is proportional to $\langle A^2 \rangle$. What we are really interested in is the fluctuation in this "signal," because these fluctuations appear as part of the system noise.

Starting from the form for $\langle A^4 \rangle$ that is analogous to Eq. [31] we have been able to follow the same logical path as shown above to prove that

$$\langle A^4 \rangle = 2 \langle A^2 \rangle^2 \quad . \qquad [47]$$

Using this result, we find that the standard deviation $\Delta(A^2)$ of $\langle A^2 \rangle$ is equal to $\langle A^2 \rangle$ itself:

$$\begin{aligned} \Delta(A^2) &= \left\langle \left( A^2 - \langle A^2 \rangle \right)^2 \right\rangle^{1/2} \\ &= \langle A^2 \rangle \end{aligned} \quad . \qquad [48]$$

These results have one particularly interesting interpretation. We see from Eq. [45] that the heterodyne signal from the rough surface is exactly zero when $\delta\gamma = 0$, that is, when the *z*-direction cosines of the probe laser and LO are the same. This condition is met when the probe and LO make the same angle with the surface normal, regardless of their relative azimuths; that is, there is a geometry of probe and LO for which the surface roughness produces no noise in the heterodyne signal.

## 2.4. System SNR

We are now ready to make an estimate of the signal-to-noise ratio for an optical heterodyne system used to detect a spherical particle on the surface of a mask blank. We will assume that the probe laser and the LO laser are single spatial mode gaussian beams having a waist at the blank surface. We will also assume that they have equal angles of incidence, so that there is no heterodyne beat signal from the rough surface. Finally, we will assume that the system noise is dominated by the shot noise from the LO.

**2.4.1. Formulas.** We will use Eq. [24] to get the power scattered by the particle. We assume that the scattering has the angular dependence shown in Eq. [26]; therefore, we can divide the total scattered power by $\pi$ to get the power scattered per unit solid angle. This is a simplification of the true scattering geometry, but for this overview we believe the results are adequate, certainly accurate to within a factor of two in the SNR. We also ascribe an effective solid angle $\Omega_{\text{eff}}$ to the heterodyne detection system, where

$$\begin{aligned} \Omega_{eff} &\equiv \frac{\lambda^2}{2\pi w_x w_y} \\ &= \frac{2\lambda^2}{\pi d^2} \end{aligned} \quad , \qquad [49]$$

where again the waist radii at the $1/e^2$ intensity points are $w_x$ in the $x$ direction and $w_y$ in the $y$ direction. In more detailed calculations not shown here we have justified this definition of $\Omega_{\text{eff}}$.

Next, using Eq. [24] we replace the product $I_S\mathcal{A}$ in Eq. [20] with the detected power $P_S$ given by

$$\begin{aligned} P_S &\equiv \frac{1000}{\pi} P_L \frac{D^6}{\lambda^4 d^2} \Omega_{eff} \\ &\approx 200\, P_L \frac{D^6}{\lambda^2 d^4} \end{aligned} \quad . \qquad [50]$$

Our final result for the SNR for the detection of a spherical "defect" of diameter $D$ is

$$\text{SNR} \approx 200 \frac{\eta P_L T}{h\nu} \frac{D^6}{\lambda^2 d^4} \quad , \qquad [51]$$

where we recall the definitions

$\eta \equiv$ photomixer quantum efficiency
$P_L \equiv$ probe laser power striking the mask
$T \equiv$ integration time
$h\nu \equiv$ energy of the laser quantum
$D \equiv$ diameter of spherical "defect"
$\lambda \equiv$ laser wavelength
$d \equiv$ diameter of laser spot at $1/e^2$ intensity.

**2.4.2. Numerical examples.** Before showing typical examples, let us define one more important parameter—the time $T_{scan}$ that it takes to scan one entire mask blank for defects. Without worrying about the details of the scan pattern, we simply estimate the scan time as about equal to the time for one measurement multiplied by the ratio of the mask area to the area of the laser spot. We put in an extra factor of two to make sure we do not miss any of the mask. Thus we define

$$T_{scan} \equiv 2T\left(\frac{W}{d}\right)^2 \quad , \qquad [52]$$

where $W$ is the diameter of the mask.

We next address the question of how large a SNR is required. Current inspection standards dictate that we must design the SNR to be large enough that there is likely to be less than one false positive detection of a defect during the scanning of an entire wafer. The number of measurements made on a wafer is approximately the ratio $T_{scan}/T$. For a typical 200-mm diameter wafer and a laser spot diameter of about 10 μm, this ratio from Eq. [52] is about $8\times10^8$. This means that the probability of a false positive on any one measurement must be less than the reciprocal of this number. If we assume that the system noise voltage is a zero-mean gaussian random variable, then we are looking for a SNR on one measurement that satisfies the relation

$$\frac{1}{\sqrt{2\pi}}\int_r^\infty e^{-t^2/2}dt \le \left(\frac{T}{T_{scan}}\right) \quad , \qquad [53]$$

where $r$ is the desired SNR. For the example numbers given above we find the requirement that SNR ≥ 6. The rule of thumb used in commercial inspection machines seems to be closer to SNR ≥ 10, so we will adopt this value also.

We now show results for two cases, an 80-nm defect and a 25-nm defect. The first case corresponds approximately to the current state-of-the-art for commercial production-line inspection tools, the second case to the 50-nm technology node shown in Table 2.

**Case 1:**

$D = 80\,\text{nm}$

$W = 200\ \text{mm}$

$\eta = 0.8$

$\lambda = 488\,\text{nm}$

$h\nu = 4.07 \times 10^{-19}\ \text{J}$

$P_L = 0.5\,\text{W}$

$d = 35\,\mu\text{m}$

$T = 0.1\,\mu\text{s}$

which gives (from Eqs. [51] and [52])

$\text{SNR} = 13$

$T_{scan} = 6.5\,\text{s}$

**Case 2:**

$D = 25\,\text{nm}$

$W = 200\ \text{mm}$

$\eta = 0.8$

$\lambda = 488\,\text{nm}$

$h\nu = 4.07 \times 10^{-19}\ \text{J}$

$P_L = 0.5\,\text{W}$

$d = 6\,\mu\text{m}$

$T = 0.1\,\mu\text{s}$

which gives (from Eqs. [51] and [52])

$\text{SNR} = 14$

$T_{scan} = 220\,\text{s}$

## 2.5. Experiments

We constructed two experiments to verify the theoretical results of the previous sections. The first of these measured how close we could come to the theoretical noise limits of optical heterodyne detection in an ideal system, but it did not involve scanning a mask blank. The second experiment is a prototype of a system to actually find defects on a mask. It consists of three major subsystems: optical, electronic and mechanical scanning. The following subsections describe both experiments, including their current status.

**2.5.1. Experimental noise limit.** Figure 4 shows a simple optical system in which we constructed a signal laser beam and a local oscillator (LO) beam, combined them on a beam splitter, and measured their beat frequency on a photodiode. The light source is a single mode (both transverse and longitudinal) linearly polarized $Ar^+$ laser that operates on several blue and green lines. The output power is hundreds of milliwatts, sufficient for providing a strong probe beam for defect scattering and a weak local oscillator beam for mixing. The laser is followed by an uncoated window that acts as a beam splitter, providing ~100:1 intensity ratio between the probe and local oscillator beams, respectively. Each beam passes through an acousto-optic frequency shifter, one centered at 60 MHz and the other at 70 MHz.

Each has a 10 MHz bandwidth. By using one or both frequency shifters, differences of in the range of 10-20 MHz and 55-75 MHz can be used depending on the electronics and/or scan speed.

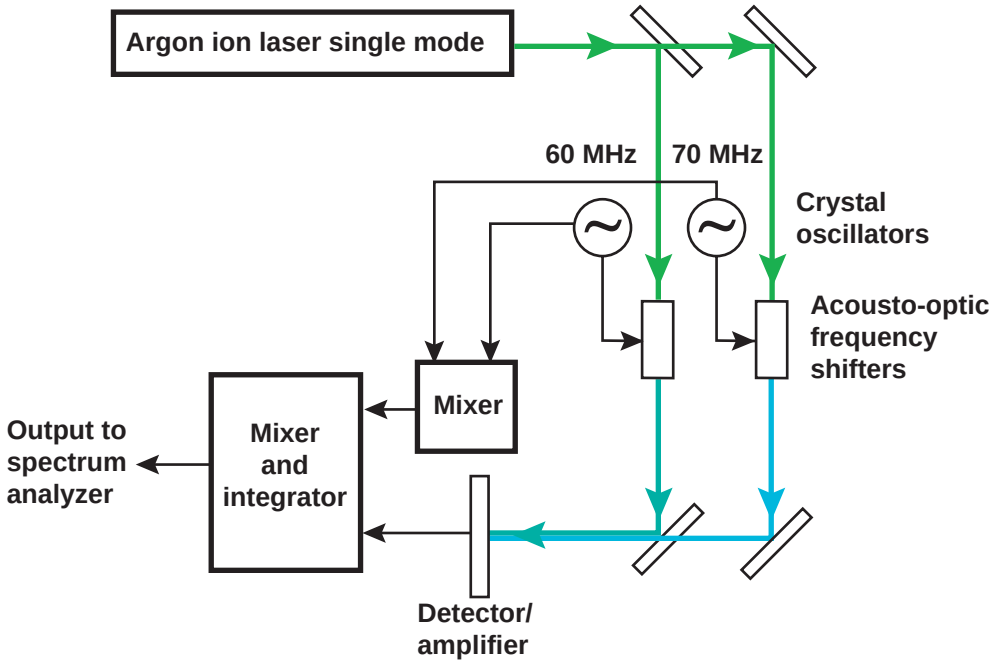


**Figure 4. Apparatus for measuring heterodyne system noise.**

We varied the signal power over a wide range including zero in order to check the calibration of the system gain. Then, at zero signal power we measured the system noise using a spectrum analyzer. Table 3 shows the results of these measurements for 10-MHz and 60-MHz beat frequencies.

At a 10-MHz beat frequency the system performed very nearly as theory predicts. The measured amplitude $A$ of the beat note was about 80% of the predicted value, presumably because the intensity profiles and wave fronts of the signal beam and the LO beam did not match perfectly. Within experimental error of about ±20% the output system noise power matched the theory. At the 60-MHz beat frequency there was substantially more noise in the photo-diode than the manufacturer specified (New Focus Model 1801 Si PIN diode). We were not able to determine the cause of this noise.

When we use Eq. [21] to calculate the minimum detectable optical power in the integration time $T = 1/2\Delta\nu$ of 1.7 μs, we get a result of 0.4 pW. This is the same result we get by extrapolating the measured signal voltage of 1.8 V for 110 μW down to the measured system noise voltage of 70 μV rms (–70 dbm). Said another way, this simple optical heterodyne system, operating in daylight, is capable of detecting 0.4 pW of laser power in a time of 1.7 μs.

**Table 3. Results of heterodyne system noise measurements at 10-MHz and 60-MHz beat frequencies.**

| | 10 MHz | 60 MHz |
|---|---|---|
| $\lambda$ | 488 nm | 488 nm |
| $R$ | 0.25 A/W | 0.25 A/W |
| $g^{\dagger}$ | $4.0 \times 10^4$ V/A | $4.0 \times 10^4$ V/A |
| $\Delta\nu$ | 0.3 MHz | 0.3 MHz |
| $P_0$ | 125 μW | 120 μW |
| $P_S$ | 110 μW | 100 μW |
| $A = 2Rg\sqrt{P_0 P_S}$ | 2.35 V | 2.2 V |
| $A$ (meas.) | 1.8±0.2 V | 0.9±0.1 V |
| Diode noise | 50±15 pW (–73 dbm) | 500±100 pW (-63 dbm) |
| Electronics noise | 3±2 pW (-85 dbm) | 4±2 pW (-84 dbm) |
| $P_N = 2qRg^2 P_0 / 50\,\Omega$ | 0.10 nW (-70 dbm) | 0.10 nW (-70 dbm) |
| $P_N$ (meas.) | 0.10±0.02 nW (-70 dbm) | (diode too noisy) |

$^{\dagger}$g is the trans-impedance amplifier gain.

**2.5.2. Prototype optical layout.** Figure 5 is a schematic drawing of our prototype system. We designed it to give maximum flexibility for wavelength selection, frequency shifting, polarization and incidence angles on the mask blank. The laser, beam splitters, and frequency shifters are the same as used above for our system noise measurement. Following the frequency shifters, each beam is focused into a single mode optical fiber, passes through a polarization adjuster, and diverges from the other end of the fiber into a focusing lens. These lenses can be positioned to illuminate the mask blank at incidence angles that optimize the signal-to-noise for an expected defect size. The lenses have a numerical aperture of 0.2, giving a focal spot diameter of 3μm. A third lens is placed in the specular reflection of the local oscillator beam to collect and focus the scattered and local oscillator beams onto a multi-mode optical fiber that terminates at the photo-diode.

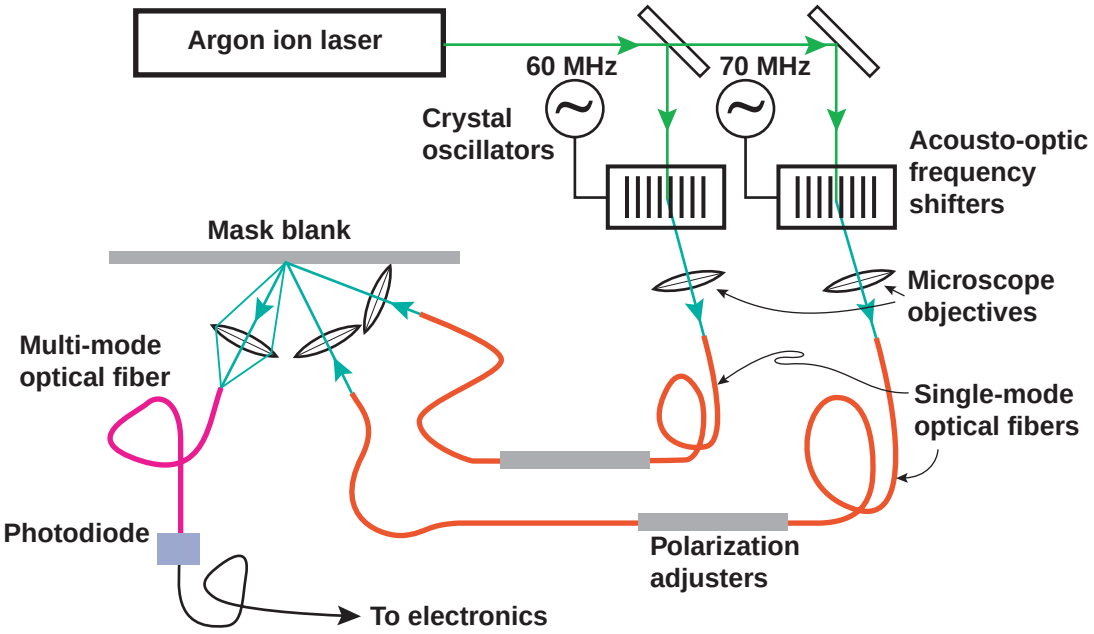


**Figure 5. Prototype optical heterodyne system for measuring wafer defects.**

**2.5.3. Electronics.** The electronics for the prototype system will consist of commercially available components having the functions shown in Fig. 3. The pair of RF mixers is available as a unit known as an I-Q modulator, used in the broadcast industry. We have in mind a simple PC-based data acquisition system to record the output of the integrator and to control the rotation/translation stage.

**2.5.4. Mechanical scanning.** For initial testing we decided to scan only a small fraction of the mask blank. The simplest method is a spiral scan used in many wafer inspection tools. This is accomplished by placing a rotary air-bearing on a linear translation stage. Figure 6 shows a CAD drawing of the assembled hardware. The air-bearing rotates at 2 rev/s to achieve a scan velocity of 1.0m/s at a radius of 80mm on the mask blank. The translation stage moves by one-half the focused spot diameter each revolution of the mask blank or about 3μm/s. This scan overlap provides some redundancy for defect detection and can also be used for correlation tests to eliminate false positives.

A specially designed vacuum chuck is fitted on the air-bearing to hold the mask blank. The surface of the chuck is polished flat and perpendicular to the rotation axis to keep the mask blank in the same plane during the scan.

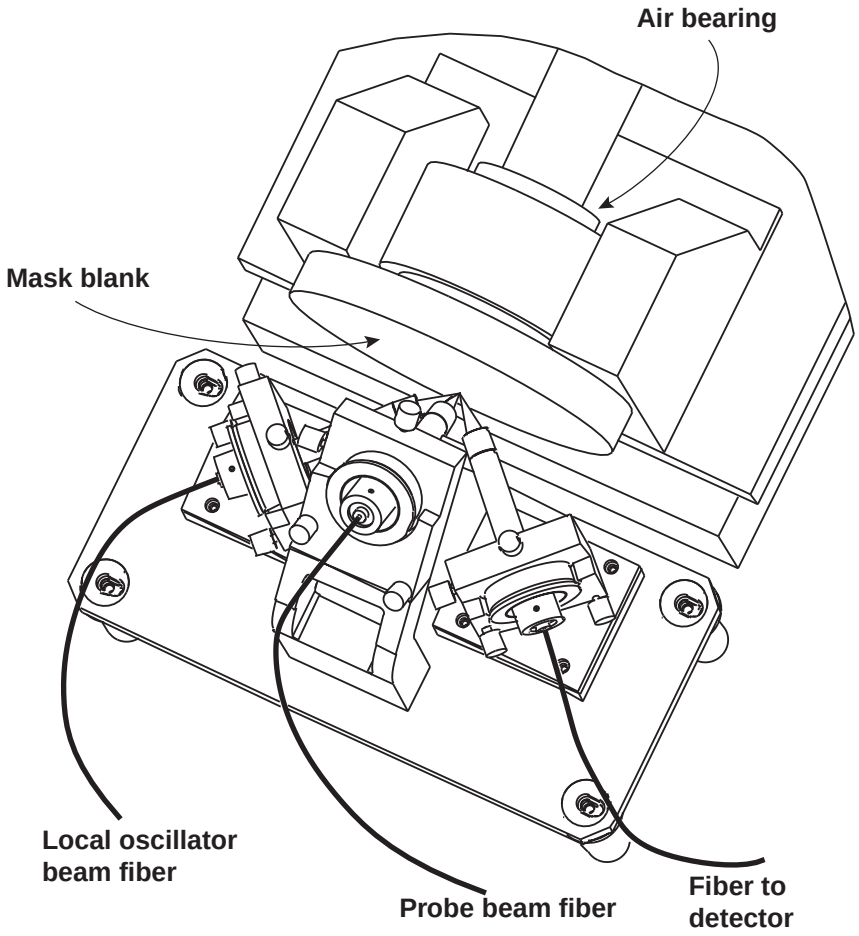


**Figure 6. CAD drawing of assembled optics and wafer translation/ rotation stage.**

**2.5.5. Programmed defects.** To compare measurements with theory it is necessary to run controlled experiments. A series of mask blanks were prepared with known defects. The substrates are high quality silicon wafers with a high-spatial-frequency roughness better than 0.2 nm rms. The defects are gold spheres that have been sorted by size and vary from 10 nm to 100 nm in diameter. Each size was randomly distributed on a single wafer.

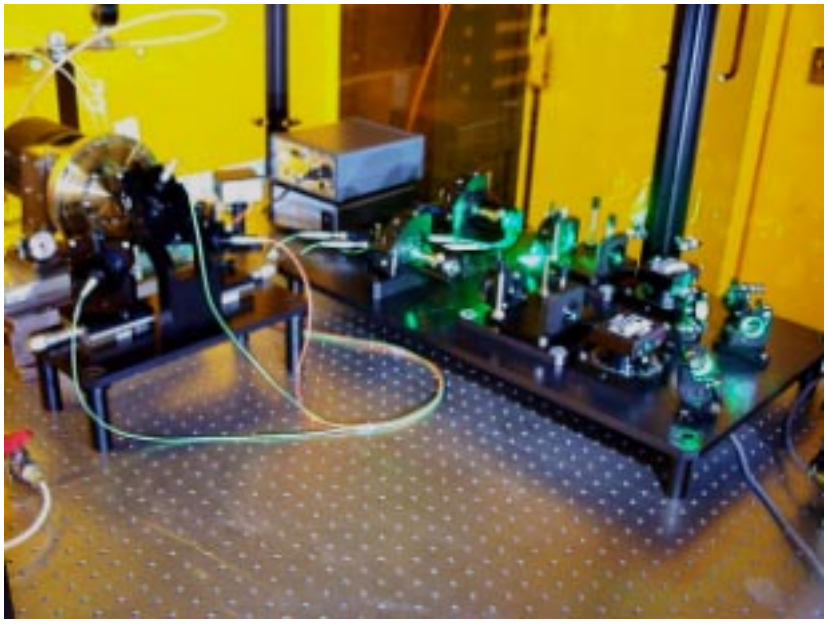

**Figure 7. Prototype system apparatus.**

## 3.0. Status of project

The theory, optical system hardware, electronics, scanning stages and defect-programmed mask blanks are in place. Figure 7 is a photograph of this prototype heterodyne system. Quantitative measurements and comparison to calculations based on the theory developed in the previous sections are the next phase of this project.

We are pursuing funding for this phase on two fronts. After we briefed Intel Corporation last fall, they made an informal commitment to fund future work for one year, starting in April 2000, at a level of approximately 1.5 FTEs. The contractual agreement will be either an add-on to an existing CRADA or an entirely separate CRADA. The final decision will depend on assignment of the intellectual property.

We have also had numerous discussions with several vice presidents at KLA/Tencor Corporation, a vendor for semiconductor inspection equipment. These discussions began with a visit by KLA to LLNL last summer. In December we submitted a plan to them for a two-year CRADA. The work would have three phases. Phase I would cover the first six months and fund 1.5 FTEs to make a theoretical comparison between heterodyne detection and direct detection. Direct detection is used in commercial wafer inspection systems, such as those manufactured by KLA, and is currently able to detect particles as small as 80 nm in diameter. Phase II would focus on the experimental comparison to theory, last nine months, and cover 2.5 FTEs. In the final phase we would develop an engineering prototype that could serve as a test bed for design of a commercial instrument. LLNL would transfer the technology to KLA during this phase, which would last nine months and support 5 FTEs.

KLA is now in the process of prioritizing their research efforts for the next fiscal year and our proposal is being weighed against other internal proposals for product development funds. We expect some feedback starting in April as their selection process narrows the field.

## References

[1] A. J. Forrester, R. A. Gudmundsen, and P. O. Johnson, *Phys. Rev.*, **99**, 1691 (1955).
[2] A. Yariv, Introduction to Optical Electronics, (Holt, Rinehart, & Winston, 1971), §10.4.
[3] W. B. Davenport, W. L. Root, An Introduction to the Theory of random Signals and Noise, (IEEE Press, 1987), §12-2.
[4] M. A. Johnson, C. H. Townes, Opt. Comm., **179**, 183 (2000).
[5] H. C. van de Hulst, Light Scattering by Small Particles, (Dover, New York, 1981).
[6] J. W. Goodman, Statistical Optics, (Wiley, 1985), 45.
[7] Davenport & Root, §8.
[8] J. C. Stover, Optical Scattering: Measurement and Analysis, (SPIE, 1995), Chapter 2.